

# Metagenomic Analysis Reveals Diverse Polyketide Synthase Gene Clusters in Microorganisms Associated with the Marine Sponge *Discodermia dissoluta*<sup>†</sup>

Andreas Schirmer,<sup>1\*</sup> Rishali Gadkari,<sup>1</sup> Christopher D. Reeves,<sup>1</sup> Fadia Ibrahim,<sup>1‡</sup> Edward F. DeLong,<sup>2§</sup> and C. Richard Hutchinson<sup>1</sup>

Kosan Biosciences, Inc., Hayward, California 94545,<sup>1</sup> and Monterey Bay Aquarium Research Institute, Moss Landing, California 95029<sup>2</sup>

Received 28 August 2004/Accepted 9 March 2005

Sponge-associated bacteria are thought to produce many novel bioactive compounds, including polyketides. PCR amplification of ketosynthase domains of type I modular polyketide synthases (PKS) from the microbial community of the marine sponge *Discodermia dissoluta* revealed great diversity and a novel group of sponge-specific PKS ketosynthase domains. Metagenomic libraries totaling more than four gigabases of bacterial genomes associated with this sponge were screened for type I modular PKS gene clusters. More than 90% of the clones in total sponge DNA libraries represented bacterial DNA inserts, and 0.7% harbored PKS genes. The majority of the PKS hybridizing clones carried small PKS clusters of one to three modules, although some clones encoded large multimodular PKSs (more than five modules). The most abundant large modular PKS appeared to be encoded by a bacterial symbiont that made up <1% of the sponge community. Sequencing of this PKS revealed 14 modules that, if expressed and active, is predicted to produce a multimethyl-branched fatty acid reminiscent of mycobacterial lipid components. Metagenomic libraries made from fractions enriched for unicellular or filamentous bacteria differed significantly, with the latter containing numerous nonribosomal peptide synthetase (NRPS) and mixed NRPS-PKS gene clusters. The filamentous bacterial community of *D. dissoluta* consists mainly of *Entotheonella* spp., an unculturable sponge-specific taxon previously implicated in the biosynthesis of bioactive peptides.

Sponges are rich sources of potentially valuable natural products, such as polyketides, nonribosomal peptides, and alkaloids (38). Natural products are the basis of >50% of the most prescribed drugs in the United States (33) and are generally obtained by fermentation of terrestrial microorganisms. However, recent attempts to further exploit terrestrial sources have met with diminishing returns, whereas the number of bioactive compounds isolated from various marine sources has risen sharply (12, 13). The majority of marine natural products have been isolated from invertebrates such as sponges, tunicates, mollusks, and bryozoans (38). Many of these compounds exhibit cytotoxic, antibacterial, antifungal, antiviral, or anti-inflammatory activity. Several bioactive compounds of marine origin are currently undergoing preclinical evaluation or have entered clinical trials, e.g., ET-743, aplidine, or kahalalide F (23), but in most cases their development is greatly hampered by limited supply.

Sponges harbor large numbers of diverse bacteria in their tissue. At present, microbial sponge communities and their genomes are poorly understood. The results of recent studies

primarily relying on fluorescence in situ hybridization (14, 16, 28) or 16S rRNA sequencing (19, 26, 56, 57) indicate that most sponges harbor uniform but phylogenetically complex microbial populations that are quite distinct from those of marine plankton or marine sediments. It has been hypothesized that the bacteria in these communities synthesize many of the associated bioactive compounds (1, 8, 52). Whereas cultivating the invertebrates for natural product synthesis is generally cost-prohibitive or impossible, the cultivation of a microorganism could secure an unlimited and inexpensive supply of a potentially important compound. However, many bacteria associated with sponges have not been cultured despite considerable effort (26, 55, 57), pointing to the need for alternative strategies. Genomic approaches aimed at isolating the biosynthetic genes and expressing them in surrogate hosts is one such alternative. This usually begins with the construction and screening of metagenomic libraries, in which large pieces of DNA isolated from mixed populations without prior cultivation are cloned and screened for targeted genes or bioactivities (45). Such an approach has resulted in the identification of a number of short biosynthetic pathways and novel biocatalysts from several soil metagenomic libraries (2, 7, 17, 54). Recently, parts of the putative pederin, bryostatin, and onnamide biosynthesis genes have been cloned from metagenomic libraries of a beetle (34, 37), a bryozoan (20, 21), and a sponge (35, 36), respectively.

We are interested in the marine sponge *Discodermia dissoluta* because a promising antitumor compound, discodermolide (18), can be isolated from it. The structure of discodermolide is consistent with biosynthesis by a bacterial type I modular

\* Corresponding author. Mailing address: Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545. Phone: (510) 731-5284. Fax: (510) 732-8401. E-mail: schirmer@kosan.com.

<sup>†</sup> Supplemental material for this article may be found at <http://aem.asm.org/>.

<sup>‡</sup> Present address: University of Nebraska, E249 Beadle Center, 19th & Vine St., Lincoln, NE 68588.

<sup>§</sup> Present address: Massachusetts Institute of Technology, Department of Civil and Environmental Engineering & Division of Biological Engineering, 48-427 MIT, 15 Vassar St., Cambridge, MA 02139.

polyketide synthase (PKS) (49), which generally consist of a set of multifunctional modules. Each module is responsible for one cycle of polyketide chain elongation. The minimal module contains a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP) domain that together catalyze a two-carbon extension of the chain. After each such extension, the resulting ketone can either be left as is or converted to a  $\beta$ -hydroxyl, a double bond, or an alkane group by stepwise processing of ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains. The substrate specificities of the ATs, the order and number of the modules, and the composition of catalytic domains within each module provide a "code" for the structure of the synthesized polyketide. Furthermore, the PKS genes for a particular polyketide are usually clustered and colinear with the order of its biosynthesis. Consequently, the structure of the polyketide product of an unknown PKS can be approximated from the sequence of its genes (49).

Although we did not identify a PKS with a modular structure consistent with biosynthesis of discodermolide, we obtained a detailed picture of the diversity of genes for biosynthesis of polyketides and nonribosomal peptides (15) in the complex microbial community of *Discodermia dissoluta*. Here we describe this diversity and a method to screen extensive sponge-derived metagenomic libraries for diverse PKS genes that should be useful for identifying target PKS genes in other sponges. Based on DNA sequence data, we also show the potential of the filamentous, sponge-specific bacteria *Entotheonella* spp. to encode nonribosomal peptide synthetases (NRPS) and mixed PKS-NRPSs, and we describe a novel, giant PKS that could be involved in the biosynthesis of multimethyl-branched fatty acids in sponge-associated bacteria.

## MATERIALS AND METHODS

**Sponge collection.** *D. dissoluta* (class, *Demospongiae*; subclass, *Tetractinomorpha*; order, *Lithistida*; family, *Theonellidae*) was collected off the coast of Curaçao, Netherlands Antilles, in early November 2001. Fourteen specimens were obtained by scuba diving at depths of 30 to 60 m and from five different locations. The sponges were immediately frozen on dry ice and shipped to us directly and were kept at  $-80^{\circ}\text{C}$ . All specimens contained discodermolide (ca. 30 to 60  $\mu\text{g}$  per g of frozen sponge), determined as described previously (18).

**Enrichment of sponge-associated bacteria.** Procedures for the separation or enrichment of sponge-associated bacteria from sponge cells have been described for several sponges (1, 42, 44) and were applied in a similar way in the present study. Frozen sponge tissue was cut into small pieces ( $<1\text{ cm}^3$ ) and placed into sterile PBSE buffer ( $1\times$  phosphate-buffered saline plus 10 mM EDTA at 10 ml/g of sponge). Collagenase (Roche Applied Science, Indianapolis, IN) was added at a final concentration of 500  $\mu\text{g}/\text{ml}$ , and the mixture was shaken on ice for 30 min. For further disruption, the sponge suspension was blended briefly. The suspension was filtered (45- $\mu\text{m}$  pore size) and centrifuged at  $500\times g$  (6 min,  $4^{\circ}\text{C}$ ). The pellet was washed twice with PBSE. Microscopic examination indicated that the pellet was enriched for filamentous bacteria, although sponge nuclei and unicellular bacteria were still present. The supernatant was centrifuged at  $8,800\times g$  (20 min,  $4^{\circ}\text{C}$ ). The pellet was resuspended in 20 ml of PBSE, and 10-ml samples were layered onto 20-ml Percoll cushions (15% Percoll in PBSE). The samples were centrifuged at  $750\times g$  (10 min,  $4^{\circ}\text{C}$ ), and the cells carefully removed from the top of the Percoll cushion and washed twice with PBSE. This fraction was highly enriched for coccoid unicellular bacteria.

**DNA isolation.** Cell pellets from bacterial enrichments were resuspended in buffer (0.5 M NaCl, 100 mM EDTA, 10 mM Tris [pH 8.0]) and treated with lysozyme (150  $\mu\text{g}/\text{ml}$  for 1 h at  $37^{\circ}\text{C}$ ) and proteinase K (0.5 mg/ml)–1% sodium dodecyl sulfate (SDS) for 2 h at  $50^{\circ}\text{C}$ . After lysis, DNA was extracted with phenol-chloroform (two to three times) and chloroform and was either concentrated by using Centricon 100 spin concentrators (Millipore, Billerica, MA) or precipitated with isopropanol. DNA was treated with DNase-free RNase and/or

further purified by cesium chloride equilibrium density gradient ultracentrifugation (43). For total sponge DNA isolation, a section of frozen sponge was placed under liquid nitrogen in a mortar, broken into pieces, and pulverized to a fine powder. The nitrogen was allowed to boil off, and small aliquots of the powder were dispersed into lysis buffer (8 M urea, 2% Sarkosyl, 350 mM NaCl, 50 mM EDTA, 50 mM Tris [pH 7.5]) at 100 ml per g of sponge tissue with slow stirring at room temperature. The lysate was extracted twice with phenol-chloroform, and the DNA was spooled from the upper aqueous phase under cold ethanol. The DNA was digested with DNase-free RNase, extracted with phenol-chloroform and ethanol precipitated.

**PCR and cloning.** In order to isolate KS gene fragments from sponge-derived DNA samples, the following degenerate PCR primers were used: degKS2F.gc (5'-GCSATGGAYCCSCARCARGSVT) and degKS5.gc (5'-GTSCCGTSC CRTGSSCYTCSAC), which were biased for G or C in some positions, and degKS2F.i (5'-GCIATGGAYCCICARCARMGIVT) and degKS5R.i (5'-GTIC CIGTICCRTGISCYTCIAC), which contained inosine in some positions. A typical reaction volume of 50  $\mu\text{l}$  contained 100 to 500 ng of sponge-derived DNA, 200 pmol of each primer, 0.2 mM deoxynucleoside triphosphate (containing 7-deaza-dGTP), 10% dimethyl sulfoxide, and 2.5 U of *Taq* DNA polymerase (Roche Applied Science). Cycle steps consisted of denaturation ( $94^{\circ}\text{C}$  for 40 s), annealing ( $55^{\circ}\text{C}$  for GC-biased primers and  $40^{\circ}\text{C}$  for inosine containing-primers for 40 s), and extension ( $72^{\circ}\text{C}$  for 75 s) for 35 to 40 cycles. In order to isolate adenylation domain gene fragments of nonribosomal peptide synthetases (NRPSs), degenerate PCR primers degNRPS-1F.i (5'-AARDSIGGIGSIGSITA YBICC) and degNRPS-4R.i (5'-CKRWAICCICKIAIYTTIAYYTG) were used; these primers were designed as described previously (51). PCR conditions were identical except for longer extension times (105 s). PCR products of approximately 700 bp (KS domains) or 1,000 bp (NRPS domains) were gel purified (QIAGEN, Inc., Valencia, CA) and cloned by using a TOPO TA cloning kit as described by the manufacturer (Invitrogen Life Technologies, Carlsbad, CA). Clones with correctly sized inserts were sequenced. Eubacterial 16S rRNA gene fragments were amplified with PCR primers Eubac-27F and 1492R as described previously (9).

**Construction and screening of metagenomic libraries.** To determine whether varying the methods of library construction would increase the diversity of the metagenomic sponge libraries, we built fosmid libraries by using randomly sheared DNA from enrichments of unicellular or filamentous bacteria and cosmid libraries by using partially digested DNA from total sponge tissue. Including the DNA isolated from total sponge tissue for library construction would capture genomes that might have been lost during cell separation. For fosmid libraries, a copy control fosmid library production kit (vector pCC1FOS) from Epicentre (Madison, WI) was used according to the manufacturer's instruction. The protocol for cosmid libraries followed that of Stratagene (La Jolla, CA) for construction of libraries in SuperCos-1 with two modifications. First, the cosmid vector was a derivative of SuperCos-1 in which the 4.2-kb AfeI fragment was self-ligated to remove the simian virus 40 and *neo* marker sequences. Second, after scaling up the optimal partial Sau3AI digests chosen from a pilot experiment, they were size fractionated on 10 to 50% sucrose gradients (35 ml in 25 mM Tris-HCl [pH 8], 25 mM EDTA, 100 mM NaCl) (43). Chloramphenicol (fosmids)- or ampicillin (cosmids)-resistant clones were grown in 384-well microtiter plates (LB plus 7% glycerol plus antibiotic) and spotted onto 22-by-22-cm nylon membranes (18,432 clones in duplicate; Amplicon Express, Pullman, WA). The membranes were hybridized overnight at  $42^{\circ}\text{C}$  in 30% formamide-containing hybridization buffer (43) and washed at low stringency with  $2\times$  SSC ( $1\times$  SSC is 0.15 M NaCl plus 0.015 M sodium citrate) plus 0.1% SDS four times for 10 min each time at room temperature and two times for 30 min each time at  $55^{\circ}\text{C}$ . DNA probes were labeled with a random-primed DNA labeling kit (Roche Applied Science) by using [ $\alpha$ - $^{32}\text{P}$ ]dCTP, and hybridization signals were visualized by using a Typhoon 9410 scanner (Amersham Biosciences, Piscataway, NJ). All Southern blot hybridizations were carried out under the same conditions.

**DNA sequencing and phylogenetic analysis.** DNA sequencing was carried out by using an ABI 3730 sequencer (Applied Biosystems, Foster City, CA). PCR amplicons cloned into pCR2.1-TOPO were sequenced with M13 universal and reverse primers, and the ends of cosmid inserts were sequenced with modified T3 and T7 primers (5'-TCTCTGTTTTGTCCGTGG and 5'-TTCCCCGAAAAG TGCCAC). After induction to high copy number, the ends of fosmid inserts were sequenced with primers recommended by the manufacturer (Epicentre). To determine the entire DNA sequence of fosmid or cosmid inserts, random shotgun libraries of 2- to 3-kb Sau3A fragments were prepared and sequenced to about sixfold coverage, and small gaps were closed by "primer walking." Sequence data were analyzed on a Macintosh computer by using Sequencher, MacVector, and the BLAST server of the National Center for Biotechnology

Information. 16S rRNA sequences were aligned by using the ARB software package (27). Preliminary phylogenetic analysis was performed by using the ARB parsimony algorithm. Aligned sequences were then exported and subsequently analyzed by using PHYLIP 3.62 (J. Felsenstein, University of Washington, Seattle). Evolutionary distances were calculated by using the F84 distance correction, a transition/transversion ratio of 2, and one rate substitution category, and empirical base frequencies were determined by using dnadist. Phylogenetic trees were generated by neighbor joining and randomized order of input species. Bootstrap datasets (1,000 each for both parsimony and distance analyses) were generated by using seqboot. Parsimony trees were generated by using dnaphars, randomizing the species input order. KS and NRPS domain sequences were aligned by using CLUSTAL W, and phylogenetic trees were created by using the neighbor-joining algorithm of CLUSTAL X (6). A total of 1,000 bootstraps were performed, and trees were visualized with the TREEVIEW program.

**Nucleotide sequence accession numbers.** Fifty-six representative 16S rRNA sequences and all KS and NRPS domain sequences shown in Fig. 2 and 4 were deposited in GenBank and given accession numbers AY89070 to AY897125 (16S rRNAs), AY897138 to AY897201 (KS domains), and AY897126 to AY897137 (NRPS domains). The two sponge-associated PKS gene clusters described have the accession numbers AY907537 (SA1\_PKS) and AY907538 (SA2\_PKS).

## RESULTS

### Morphological diversity of the sponge microbial community.

We surveyed the morphology of the microorganisms associated with *D. dissoluta* by microscopy of sponge thin sections (for photomicrographs, see Fig. S1 in the supplemental material). By far the most abundant morphological group was represented by small 0.5- to 1.5- $\mu$ m coccus-shaped bacteria. Two different types of less abundant filamentous bacteria were also observed: the first type did not show autofluorescence and was found throughout the sponge tissue (its affiliation will be discussed below). The second group was autofluorescent and found exclusively attached to the pinacoderm and was most likely a cyanobacterial population. In addition to sponge cells another microbial type was found preferably in the mesohyl. These microorganisms were typically 5 to 10  $\mu$ m long and, according to their shape and fluorescent behavior, probably represent diatoms of the Nitzschoid group (25). After DAPI (4',6'-diamidino-2-phenylindole) staining of sponge thin sections, digitized images were analyzed with ImagePro software (data not shown). We estimate on this basis that *D. dissoluta* harbors roughly 100 to 200 bacterial cells for each sponge nucleus.

### Phylogenetic diversity of the sponge-microbe community.

The phylogenetic diversity of bacteria associated with *D. dissoluta* was surveyed from 160 eubacterial 16S rRNA sequences cloned from total sponge DNA or DNA of bacterial enrichments. The majority of 16S rRNA sequences in the unicellular and filamentous fraction were related to acidobacteria ( $n = 22$ ; 45%) and  $\delta$ -proteobacteria ( $n = 23$ ; 45%), respectively. Members of other phylogenetic groups that were identified multiple times were *Chloroflexi* ( $n = 21$ ),  $\gamma$ -Proteobacteria ( $n = 17$ ),  $\alpha$ -Proteobacteria ( $n = 8$ ), *Thermus/Deinococcus* group ( $n = 9$ ), *Spirochaetes* ( $n = 5$ ), and *Actinobacteria* ( $n = 5$ ). The presence of eubacteria from most of these groups has also been reported for other sponges (19). Not surprisingly, 76% of all eubacterial 16S rRNA sequences were most similar (91 to 99% identical) to other sponge-derived 16S rRNAs (see Table S1 in the supplemental material). Interestingly, all of the  $\delta$ -proteobacteria, which make up the majority of the filamentous fraction, were closely related and found to be affiliated with *Entotheonella*, a proposed novel sponge-specific taxon (46). Figure 1 shows the

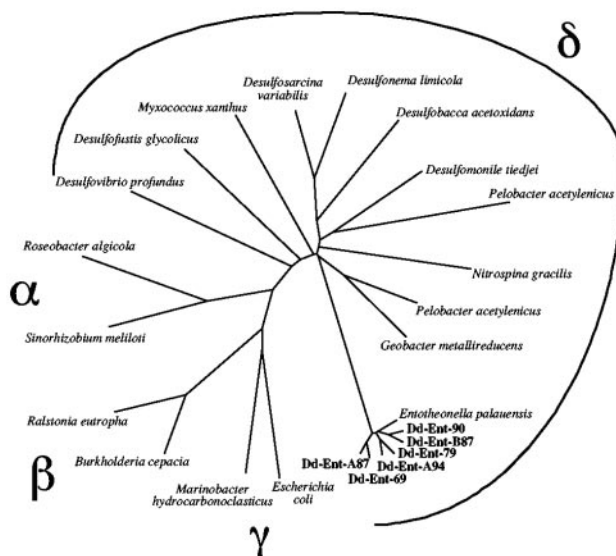


FIG. 1. Phylogenetic tree of 16S rRNA sequences that make up the majority of the filamentous bacteria associated with *D. dissoluta* (bold) showing their close affiliation with “*Candidatus Entotheonella palauensis*” and classification within the proteobacteria.

phylogenetic relationship of the six different strains identified (96 to 98% identical to each other over 1480 bp of 16S rRNA) with the unculturable “type strain” “*Candidatus Entotheonella palauensis*” (ca. 97% identical over 1,335 bp of 16S rRNA). Moreover, the morphology of the filamentous bacteria found throughout the *Discodermia* tissue matches that of *E. palauensis* (46).

**Distribution of PKS KS domains in sponge-derived genomic DNA.** Using degenerate KS primers hybridizing to conserved regions within the KS domain (39), we amplified and analyzed 256 KS domain DNA sequences (ca. 680 bp) from the metagenome of *D. dissoluta*. Based on the criterion that sequences that were >95% identical might have been amplified from identical genes (we have seen up to 5% of PCR-induced errors in this kind of degenerate PCR), 85 KS sequences were considered to be different. A total of 86% of KS genes isolated from the unicellular fraction had a relatively high G+C content (65 to 75%), whereas 87% of the KS genes isolated from the filamentous fraction had a lower G+C content (50 to 65%). All KS domain sequences were analyzed by BLAST, and the deduced amino acid sequences of the most diverse KS domains (a total of 44 with >20% DNA sequence divergence) were placed in a phylogenetic tree (Fig. 2). The tree shows the high diversity of KS domains from microorganisms associated with *D. dissoluta*. For most of these KS sequences, the percent identity to their most homologous database sequences was fairly low, ranging from 49 to 74%. Seventeen KS domains were most homologous to cyanobacterial or myxobacterial type I PKS genes, and four were predicted to come from “trans-AT”-type PKS gene clusters such as pederin (34), in which AT domains are encoded by discrete genes outside of PKS modules (5). Another three KS domains resembled those following a NRPS module in hybrid NRPS/PKS clusters (32), such as MtaD (48). Eighteen KS domains (cluster at bottom in Fig. 2)





FIG. 2. Phylogenetic analysis of type I PKS KS domains from sponge-associated microorganisms. Bootstrap values of >500 calculated from 1,000 bootstrap trees (neighborhood-joining algorithm) are indicated at the nodes. The bar indicates 10% sequence divergence. The tree was rooted with *Escherichia coli* KasI (6573501) as an outgroup (not shown). Sponge-derived KS domains are in boldface and have the following letter code: sp, from genomic DNA; fos/cos, from fosmid/cosmid DNA (highlighted by asterisks); T, from total sponge; and F or U, from enrichment of filamentous or unicellular bacteria, respectively. Known and hypothetical KS domains from GenBank used in this tree, which are given either with their gene symbol or accession number, represent the closest homologs of the sponge-derived sequences (as judged by BLAST) and selected more distant relatives. Thsw-sp, KS domains isolated from the sponge *Theonella swinhoei*.

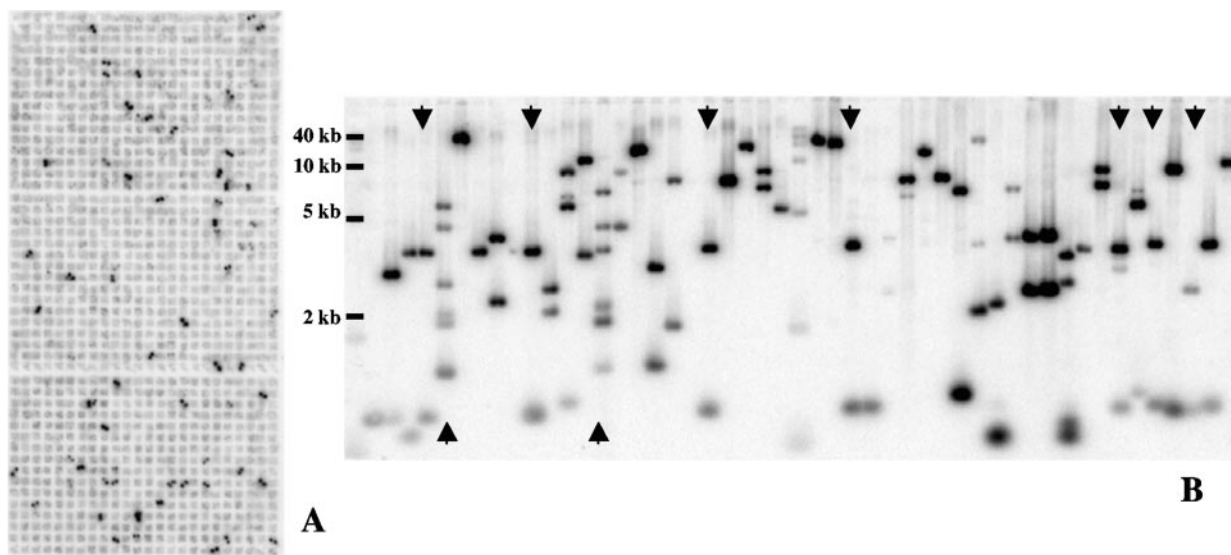


FIG. 3. Hybridization experiments at low stringency using a KS probe pool. (A) Colony hybridization of a section of a macroarrayed total sponge cosmid library (9,216 clones). (B) Southern blot hybridization of a BamHI digest of randomly chosen KS-positive fosmids from the unicellular library. Arrows at the top and bottom depict fosmids containing an apparently abundant PKS gene cluster and fosmids containing large multimodular PKSs, respectively.

stand out in that they are most closely related to each other or to KS domains derived from the sponge *T. swinhoei* (63 to 95% identity) (35). These sequences form a distinct group that consists only of KS domains derived from sponge-associated microorganisms. Their phylogenetic relationship is supported by high bootstrap values.

#### Phylogenetic distribution in metagenomic sponge libraries.

To assess the range of prokaryotic DNA inserts in the metagenomic libraries made from either enriched bacterial fractions or total sponge, we sequenced the ends (500 to 600 bp) of ca. 340 random fosmids or cosmids. All sequences were searched against the NCBI nonredundant databases and assigned to the *Prokarya* (best BLASTX expectation value of  $\leq 10^{-4}$  to a prokaryotic entry) or *Eukarya* (best BLASTX expectation value  $\leq 10^{-2}$  to a eukaryotic entry) or considered not assignable (not meeting cutoff values or comparable similarity to prokaryotic and eukaryotic entries). Based on this analysis, at least 46 and 82% of fosmid inserts from the filamentous and unicellular libraries, respectively, encode prokaryotic proteins or open reading frames (ORFs). The number of clearly assigned eukaryotic fosmid inserts was low in both bacterial libraries and yet significantly higher in the filamentous library (20% compared to 3%). This finding correlates with the higher number of sponge cells observed in the enriched filamentous fraction compared to the enriched unicellular fraction. At least 87% of cosmid and 93% of fosmid inserts from two different total sponge libraries encoded prokaryotic and only 1 to 2% eukaryotic proteins or ORFs, although the construction of these libraries did not include any physical enrichment of sponge-associated bacteria.

**Distribution of type I PKS gene clusters in metagenomic sponge libraries.** We used a probe pool consisting of 44 diverse KS domain sequences isolated from sponge-derived genomic DNAs (see above and Fig. 2) to screen high-density filters of macro-arrayed clone libraries at low stringency (an example is

shown in Fig. 3A). Not only do these macro arrays allow for high-throughput screening, they also reduce the number of false positives because each clone is spotted in duplicate, thus facilitating the identification of low-homology targets. To capture the genomes of the less-abundant sponge-associated microbes, as well as the abundant ones, we arrayed about 155,000 clones with an average insert size of 35 kb. Allowing for the proportion of bacterial inserts in each library, this screen covered more than 4 Gb, the equivalent of more than 1,000 sponge-associated bacterial genomes of average size (Table 1). The frequency of PKS positive clones was about one in 140 (0.7%) for the unicellular and total sponge library and one in 200 (0.5%) for the filamentous library, yielding 1,025 PKS gene-containing fosmids or cosmids [from here on referred to as PKS f(c)osmids]. Based on end sequencing of 467 PKS f(c)osmids, ca. 80% of the PKS f(c)osmids from the unicellular and total sponge library had a G+C content of >60%. In contrast, the G+C content of 65% of PKS fosmids from the filamentous library was <60%, indicating that this library contained a very different set of PKS gene clusters.

Southern blot hybridization of about 640 PKS f(c)osmids with the same KS probe pool revealed that certain PKS clusters were isolated multiple times (having identical KS hybridization patterns, examples are shown in Fig. 3B). The most abundant PKS cluster in the unicellular library represented ca. 10% of the unicellular PKS fosmid collection. Likewise, a distinct PKS cluster made up 7% of the PKS fosmids in the filamentous library. These findings suggest that each library was redundant for certain genomes. Most clones showed only one to three hybridizing bands (examples are shown in Fig. 3B), suggesting that the majority of cloned PKS gene clusters consisted of only one or a few modules. A high frequency of small PKS gene clusters was confirmed by sequencing the ends of PKS f(c)osmids (Table 1): 50 to 68% did not have PKS gene sequence at either end, indicating that these PKSs were smaller than 35 kb

TABLE 1. Distribution of modular PKS gene clusters in metagenomic libraries of *D. dissoluta*

Parameter	Library		
	Unicellular bacterial enrichment	Filamentous bacterial enrichment	Total sponge tissue
No. of f(c)osmids screened	64,000	36,000	55,000
Equivalent size (Mb) of prokaryotic DNA <sup>a</sup>	1,840	580	1,680
No. of equivalent of prokaryotic genomes <sup>b</sup>	460	145	420
No. of PKS-positive f(c)osmids <sup>c</sup> (%)	465 (0.7)	185 (0.5)	375 (0.7)
No. of PKS f(c)osmids end sequenced	145	151	171
% F(c)osmids with two PKS ends	2	1	4
% F(c)osmids with one PKS end	29	49	40
% F(c)osmids with no PKS end	68	50	56
% PKS f(c)osmids with NRPS end	0	13	1
% <i>trans</i> -AT type PKS f(c)osmids	0	11	1

<sup>a</sup> That is, the minimum number of prokaryotic f(c)osmids [based on end sequencing of random f(c)osmids] multiplied by an average insert size of 35 kb.

<sup>b</sup> That is, the equivalent of prokaryotic DNA divided by an average prokaryotic genome size of 4 Mb.

<sup>c</sup> As determined by hybridization with sponge-derived KS probes.

or five modules (PKS modules have an average size of 5 kb). Furthermore, only 1 to 4% of the PKS f(c)osmids had PKS gene sequence at both ends, a finding indicative of large multimodular PKSs with more than five modules. To see whether the KS segments identified before library construction were present in the metagenomic libraries, we amplified and analyzed the KS domains from approximately 40 PKS f(c)osmids. Fifty-seven different KS domains were isolated, but only 15% were found to be 90 to 100% identical to KS domains amplified from sponge-derived DNAs. Twenty of the most diverse KS domains, fifteen of which belonged to the sponge-specific KS domain group, were included in the phylogenetic KS tree in Fig. 2, showing their divergence from the KS probes.

**Distribution of mixed PKS-NRPS and NRPS gene clusters in metagenomic sponge libraries.** Another striking difference between the filamentous and unicellular libraries was that 13% of the PKS fosmids from the filamentous library had end sequences with high homology to NRPS genes, whereas no NRPS-related end sequences were found on any PKS fosmids from the unicellular library (Table 1). In agreement with this, degenerate primers specific for NRPS adenylation (A)-domains amplified correctly sized DNA fragments from some PKS fosmids in the filamentous library but from none in the unicellular library. These results encouraged us to examine the frequencies of mixed PKS-NRPS and NRPS gene clusters in the bacterial sponge libraries. NRPS A-domain amplicons were cloned and sequenced from several mixed PKS-NRPS fosmids and from total sponge DNA. A total of 12 different NRPS A-domain sequences were identified, which were only 42 to 61% identical to their closest homologs in GenBank, mostly cyanobacterial NRPS A-domains (a phylogenetic tree is shown in Fig. 4). These A-domain amplicons were used as probes to screen parts of the filamentous and unicellular libraries at low stringency. About 1% of the fosmids in the filamentous library hybridized strongly to these probes, including 33% of the PKS fosmids, which indicates that these fosmids probably contain mixed PKS-NRPS gene clusters. In contrast, no NRPS-positive fosmids could be detected in the unicellular library (see Fig. S2 in the supplemental material for hybridization data).

**Distribution of “*trans*-AT”-type PKS gene clusters in metagenomic sponge libraries.** It was shown recently that the KS

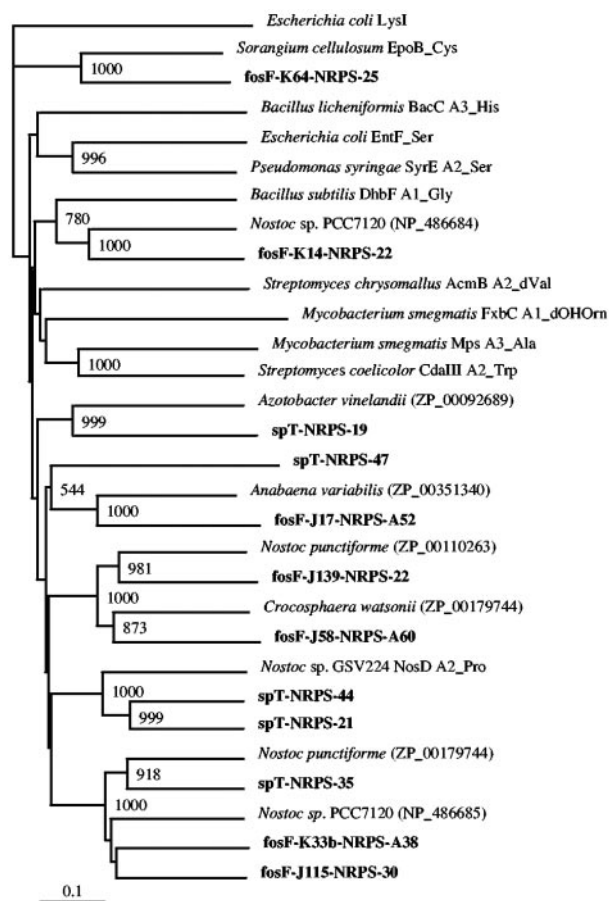


FIG. 4. Phylogenetic analysis of NRPS adenylation (A) domains from sponge-associated microorganisms. Bootstrap values of >500 calculated from 1,000 bootstrap trees (neighborhood-joining algorithm) are indicated at the nodes. The bar indicated 10% sequence divergence. The tree was rooted with *E. coli* LysI (P40976) as an outgroup. Sponge-derived NRPS A-domains are in boldface and have the same letter code as in Fig. 2. Known and hypothetical NRPS A-domains from GenBank used in this tree, which are given either with their gene symbol and amino acid specificity or accession number, represent the closest homologs of the sponge-derived sequences (as judged by BLAST) and selected more distant relatives.



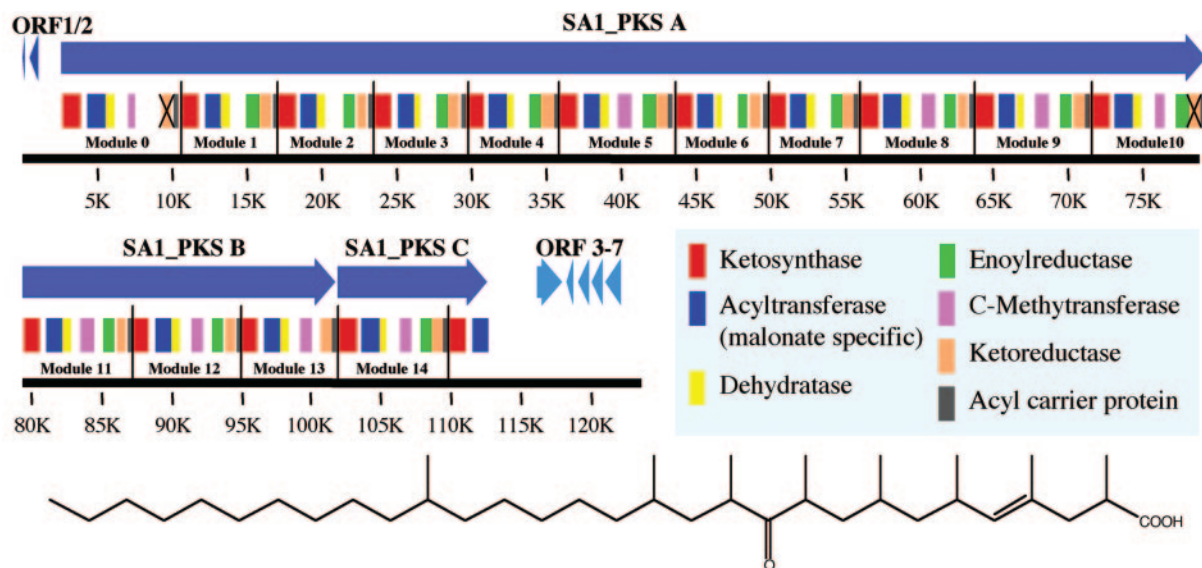


FIG. 5. Organization of the most abundant large multimodular PKS gene cluster in metagenomic libraries from *D. dissoluta*. The module and color-coded domain structure of the PKS genes is shown at the top, and the predicted polyketide product at the bottom. Crossed-out domains within the PKS are predicted to be inactive. The closest homologs and amino acid identity of ORFs flanking the PKS are as follows: ORF1, 34% to *Stigmatella aurantiaca* phosphopantetheinyl transferase MtaA (AAF19809); ORF2, 46% to *Streptomyces avermitilis* SIMX4 (AAK06800); ORF3, 60% to *Sphingomonas* sp. strain KA1 insertion sequence ATP-binding protein (BAC56754); ORF4, 57% to *Magnetococcus* sp. strain MC-1 transposase (ZP\_00291288); ORF5 to ORF7, 42 to 49% to *Mesorizhobium loti* intergrase/recombinase ORFs NP10678 to NP10680.

domains of the unusual “*trans*-AT” type PKSs can be distinguished from those of the more usual “*cis*-AT”-type PKSs by a phylogenetic approach (35). Thus, we searched the end sequences of the PKS f(c)osmids for sequences with highest homology to “*trans*-AT” type PKS genes. As with NRPS-containing gene clusters, Table 1 shows there were no “*trans*-AT” type PKS genes in the unicellular library, whereas ca. 11% of the end sequences of PKS fosmids from the filamentous library appeared to be of the “*trans*-AT” type. The most abundant PKS gene cluster in the filamentous library belonged to this type, too. We sequenced one fosmid of this cluster in order (i) to confirm that it truly was a “*trans*-AT” type PKS and (ii) to obtain an insight into the nature of this abundant PKS gene cluster in the filamentous library, which could be encoded by *Entotheonella* ssp. The 31-kb DNA sequence had an overall G+C content of 54.8% and contained the last three modules of a modular PKS terminated by a thioesterase. As expected, none of the modules contained ATs. Adjacent to the PKS we identified a pathway that is thought to be responsible for the incorporation of exomethyl or exomethylene groups at the  $\beta$ -keto position of polyketides (11, 37). A more detailed description of this PKS cluster (SA2\_PKS) can be found in Fig. S3 and Table S2 in the supplemental material.

**Distribution of large multimodular PKS gene clusters in metagenomic sponge libraries.** The modular PKSs that synthesize bioactive polyketides are usually large and comprise multiple modules; for example, the erythromycin PKS contains six, the avermectin PKS 12 and the nystatin PKS 18 modules (49). The discodermolide PKS is predicted to consist of 10 or 11 modules. Of approximately 1,000 PKS f(c)osmids, 13 (1.3%) contained only PKS sequence (~35 kb) and thus were part of large PKSs with more than five modules. Restriction mapping and Southern blot hybridization revealed that these PKS

f(c)osmids belonged to four distinct PKS gene clusters, and additional overlapping PKS f(c)osmids containing the start and/or end of some of these clusters were identified. The most abundant large multimodular PKS gene cluster was covered by 10 overlapping f(c)osmids spanning 125 kb (Fig. 5). Its full DNA sequence was determined from four f(c)osmids. Overlapping segments 1 to 14 kb long of nine f(c)osmids had identical sequences and therefore come from the same genome and not a set of closely related genomes, as is possible in metagenomic libraries (53). Because these f(c)osmids gave strong hybridization signals in the initial screen and our analysis of all PKS f(c)osmids by Southern blot hybridization and/or end sequencing was very rigorous, it seems unlikely that we missed a significant number of additional clones of this PKS gene cluster. Therefore, we can estimate the relative abundance of the microorganism that encodes this PKS: seven cosmids (representing 250 kb) were identified in the total sponge library, which corresponds to a twofold coverage of this 125-kb locus. Since the total sponge library contained an equivalent of about 400 average bacterial genomes (Table 1), this PKS gene cluster would be encoded by a bacterium that makes up ca. 0.5% (2 in 400) of the total sponge microbial community (assuming no significant cloning bias).

**Characterization of a large multimodular PKS gene cluster from a sponge-associated bacterium.** Two ORFs upstream of this PKS, homologs of phosphopantetheinyl transferases and SimX4 (50), respectively, are possibly involved in the essential posttranslational activation (phosphopantetheinylation) of ACP domains within the adjacent PKS. The polyketide synthase consists of three PKS ORFs with an overall G+C content of 68.9% spanning 110 kb (Fig. 5); each ORF is preceded by a potential ribosome-binding site rich in A and G. The first PKS ORF codes for a remarkably large protein of 25,572 amino

acids with a predicted molecular mass of  $\sim 2.7$  MDa. The PKS contains a starter module and 14 complete and 1 incomplete extender modules. The last module is truncated by what appears to be a transposon insertion. Although this transposition event could render the PKS inactive, it does not necessarily have to (see Discussion). The starter module consists of a KS<sup>S</sup> domain (active-site cysteine replaced by serine), which is similar to starter modules of polyene polyketide synthases, e.g., nystatin PKS (3). The KS domains were only 55 to 59% identical to their closest KS domain homologs in the database (hypothetical PKS from *Mycobacterium avium*, accession NP\_961164) and belonged to the sponge-specific KS domain group (Fig. 2). All modules contain ATs, which appear to be specific for the incorporation of malonyl extender units. The most remarkable features of this large PKS gene cluster are the presence of (i) a complete set of reductive domains (KR, DH, and ER) in all except one module, which lacks the ER, and (ii) C-methyltransferase domains in 8 of the 14 modules (Fig. 5). Upon close inspection of amino acid sequences of all domains, only two KR domains (Fig. 5) showed signs of degeneration of sequence homology, suggesting that they are inactive. All other domains seemed well conserved, with an interesting variation: in five KRs the putative catalytic tyrosine (40) is replaced by a histidine. It is very likely that this particular change in so many KR domains leaves these otherwise normal KRs active. According to the well-established rules for polyketide biosynthesis (49) and assuming the loading of acetate by the starter module, the putative product of this unusual PKS would not be a complex polyketide but rather a multimethyl-branched C<sub>30</sub> fatty acid, as depicted in Fig. 5.

## DISCUSSION

We have vigorously applied a method that should be generally applicable to identify PKS genes cluster in sponge-associated microbial communities. We began by amplifying KS domains from metagenomic DNA samples of *D. dissoluta*. These KS amplicons showed high diversity and revealed a group of related KS domains thus far only found in sponge-associated microorganisms. We selected and pooled 44 of these KS amplicons and used them as probes to screen macroarrayed, metagenomic sponge libraries at low stringency. Interestingly, our analysis of a subset of PKS gene-containing f(c)osmids revealed that the majority of KS domains that we cloned from the libraries differed significantly from those used as probes. Thus, the KS amplicons represented only a small fraction of the PKS gene clusters within the sponge-microbial metagenome but nevertheless were sufficient to identify a diverse set of PKS genes. However, it is possible that our screen failed to identify KS domains with even greater diversity (i.e., lower homology to our probes).

The metagenomic libraries made from total sponge tissue surprisingly contained about 90% prokaryotic inserts. Only a few sponge genome sizes have been estimated and range from as little as 60 Mb (31) to as much as 1,700 Mb (22). If there are actually 100 to 200 bacterial cells per sponge nucleus as we estimated, the genome of *D. dissoluta* would have to be at the small end of the size range in order to give rise to metagenomic libraries with such high proportions of prokaryotic DNA inserts. Alternatively, sponge DNA might not clone as well as

bacterial DNA, biasing the metagenomic libraries toward bacterial inserts. There has been at least one report that, for unknown reasons, genomic sponge DNA was difficult to digest with restriction endonucleases (47). Interestingly, we recently made a fosmid library from total genomic DNA of another marine sponge and found that this metagenomic library also contains >90% bacterial inserts (unpublished results). At least for the two sponge metagenomic libraries we made, it appears that, with regard to the proportion of bacterial DNA inserts, the enrichment of bacteria does not provide any advantage over the use of total sponge tissue.

Metagenomic libraries made from total sponge and enriched unicellular bacteria showed several similarities, e.g., similar G+C profiles and PKS frequencies, which agrees with our microscopic observation that the vast majority of sponge-associated bacteria in *D. dissoluta* are unicellular. The apparent lack of NRPS genes in the unicellular library indicates either that such gene clusters are rare in the unicellular bacteria associated with *D. dissoluta* or that the NRPS genes encoded by these unicellular bacteria are significantly divergent from the primers and probes used for their detection.

The metagenomic library made from enriched filamentous bacteria was notably different from the other libraries; for example, the G+C content was significantly lower, and it contained numerous NRPS and mixed NRPS-PKS gene clusters. Nonribosomal peptides or mixed polyketides and peptides have not been reported for *D. dissoluta*, but several such compounds exhibiting bioactivity have been isolated from other *Discodermia* species, e.g., discodermins from *D. kiiensis* (29) or calyculins from *D. calyx* (24). Since the filamentous bacterial fraction appears to be dominated by *Entotheonella* spp., it is tempting to speculate that the abundance of NRPS genes might reflect the high abundance of *Entotheonella* genomes in this library. Interestingly, *Entotheonella* spp. have been found thus far only in sponges of the family *Theonellidae*, where they are thought to produce dicyclic glycopeptides such as theopalauamide (1, 46). Based on microscopy, the filamentous bacteria are only a minor part of the sponge-associated microbial community of *D. dissoluta*. The example shown here demonstrates the importance of enriching for subpopulations whenever possible, since this yielded a metagenomic library that was enriched in otherwise rarely found NRPS and mixed NRPS-PKS gene clusters.

We successfully isolated, assembled, and characterized a >100-kb PKS gene cluster encoded by a microbe that appears to make up <1% of the sponge bacterial community. If active, this PKS is predicted to synthesize a multimethyl-branched C<sub>30</sub> fatty acid. Interestingly, two of the other three large PKS gene clusters identified in the sponge metagenomic libraries contained PKSs with very similar structures (partially sequenced; data not shown), suggesting the existence of a family of PKSs for multimethyl-branched fatty acids in sponge-associated bacteria. Very similar multimethyl-branched fatty acids are known as part of complex glycolipids in mycobacteria, e.g., mycosesic and phthioceranic acids (30). They are formed by monomodular, multifunctional PKS enzymes, MAS and Pks2, which act like iterative vertebrate fatty acid synthases. In contrast to those of the vertebrate enzymes, however, the AT and KS domains of these enzymes are predicted to show selectivity for methylmalonyl coenzyme A (CoA) over malonyl-CoA, produc-



ing multimethyl-branched acids from *n*-acyl primers. If the PKS from the *D. dissoluta* metagenome described here indeed produces such a multimethyl-branched fatty acids, it would do so by a different and novel mechanism: a large multimodular PKS would synthesize these fatty acids de novo from malonyl-CoA, placing methyl branches at specific positions by the action of C-methyltransferase domains within its modules. It is conceivable that these fatty acids are directly transferred to lipid components and do not require a thioesterase for release, which is the case for mycosideric and phthioceranic acids (30). If so, the transposon insertion in module 15 could still leave an active PKS enzyme with 14 modules. The fatty acid composition of *D. dissoluta* has not been studied. However, mono-, di-, tri-, and tetramethyl-branched, as well as polyunsaturated, fatty acids have been identified in marine and freshwater sponges (4, 10, 41). It has been hypothesized that these unusual fatty acids come from the sponge-associated bacteria. Our findings support this idea and provide a first clue into the biosynthesis of multimethyl-branched fatty acids in sponges.

It is difficult to determine whether a metagenomic sponge library is truly representative of the sponge-microbial community. Bacterial species may be lost during enrichment procedures. For this reason, the use of total sponge tissue to prepare metagenomic libraries seems to be the least biased. However, differential cell lysis and DNA stability, as well as cloning biases during library construction, can lead to genome misrepresentation. Since these parameters are difficult to measure, it seems advisable to build in as many variations as possible when one considers exploring metagenomic sponge libraries. At present, it cannot be known whether the abundant PKS gene clusters observed come from truly abundant species or rather easily accessible and clonable sponge-associated bacterial genomes. Much more work is needed to address this issue. Our observation that (i) none of the KS domains identified from sponge-associated bacteria were related to the common "cis-AT" type polyketide synthases of actinobacteria and (ii) the clear majority of PKS gene clusters isolated from the sponge metagenome were apparently rather small (one to three modules) might indicate that the genomes of some polyketide producing sponge-associated bacteria could have been underrepresented for the aforementioned reasons or are so rare that they were difficult to access in these metagenomic libraries. This could explain why none of the isolated multimodular PKS gene clusters appear to be responsible for the biosynthesis of discodermolide, the most abundant polyketide in this sponge.

#### ACKNOWLEDGMENTS

We thank Dan Morse (University of California, Santa Barbara) and David Myles for obtaining *Discodermia dissoluta*, Steve Bernstein (QualTek Molecular Laboratories, Santa Barbara, CA) for help with microscopy on sponge thin sections, and Ralph Reid for suggestions regarding the sequence analyses. We also thank Christian Siebel, David Hopwood, and Leonard Katz for critical comments during preparation of the manuscript.

This study was supported in part by a Small Business Innovative Research grant from the National Institutes of Health (1R43CA97889-01).

#### REFERENCES

- Bewley, C. A., N. D. Holland, and D. J. Faulkner. 1996. Two classes of metabolites from *Theonella swinhoei* are localized in distinct populations of bacterial symbionts. *Experientia* **52**:716–722.
- Brady, S. F., C. J. Chao, J. Handelsman, and J. Clardy. 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* **3**:1981–1984.
- Brautaset, T., O. N. Sekurova, H. Sletta, T. E. Ellingsen, A. R. Strom, S. Valla, and S. B. Zotchev. 2000. Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. *Chem. Biol.* **7**:395–403.
- Carballeira, N. M., and M. Pagan. 2000. Identification and total synthesis of a novel dimethylated fatty acid from the Caribbean sponge *Calyx podatypa*. *J. Nat. Prod.* **63**:666–669.
- Cheng, Y. Q., G. L. Tang, and B. Shen. 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. USA* **100**:3149–3154.
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson. 2003. Multiple sequence alignment with the CLUSTAL series of programs. *Nucleic Acids Res.* **31**:3497–3500.
- Courtois, S., C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helyncck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerinneau, P. Jeannin, P. Simonet, and J. L. Pernodet. 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* **69**:49–55.
- Davidson, S. K., S. W. Allen, G. E. Lim, C. M. Anderson, and M. G. Haygood. 2001. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "*Candidatus Endobugula serula*" of the bryozoan *Bugula neritina*. *Appl. Environ. Microbiol.* **67**:4531–4537.
- DeLong, E. F. 1992. Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA* **89**:5685–5689.
- Dembitsky, V. M., T. Rezanka, and M. Srebnik. 2003. Lipid compounds of freshwater sponges: family Spongillidae, class Demospongiae. *Chem. Phys. Lipids* **123**:117–155.
- Edwards, D. J., B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, and W. H. Gerwick. 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**:817–833.
- Faulkner, D. J. 2002. Marine natural products. *Nat. Prod. Rep.* **19**:1–48.
- Faulkner, D. J. 2000. Marine pharmacology. *Antonie Leeuwenhoek* **77**:135–145.
- Fieseler, L., M. Horn, M. Wagner, and U. Hentschel. 2004. Discovery of the novel candidate phylum "Poribacteria" in marine sponges. *Appl. Environ. Microbiol.* **70**:3724–3732.
- Finking, R., and M. A. Marahiel. 2004. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**:453–488.
- Friedrich, A. B., H. Merkert, T. Fendert, and J. Hacker. 1999. Microbial diversity in the marine sponge *Aplysina cavernicola* (formerly *Verongia cavernicola*) analyzed by fluorescence in situ hybridization. *Mar. Biol.* **134**:461–470.
- Gillespie, D. E., S. F. Brady, A. D. Bettermann, N. P. Cianciotto, M. R. Liles, M. R. Rondon, J. Clardy, R. M. Goodman, and J. Handelsman. 2002. Isolation of antibiotics turbinomycin A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* **68**:4301–4306.
- Gunsakera, S. P., M. Sunasekara, R. E. Longley, and G. K. Schulte. 1990. Discodermolide: a new bioactive polyhydroxylated lactone from the marine sponge *Discodermia dissoluta*. *J. Org. Chem.* **55**:4912–4916.
- Hentschel, U., J. Hopke, M. Horn, A. B. Friedrich, M. Wagner, J. Hacker, and B. S. Moore. 2002. Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.* **68**:4431–4440.
- Hildebrand, M., L. E. Waggoner, G. E. Lim, K. H. Sharp, C. P. Ridley, and M. G. Haygood. 2004. Approaches to identify, clone, and express symbiont bioactive metabolite genes. *Nat. Prod. Rep.* **21**:122–142.
- Hildebrand, M., L. E. Waggoner, H. Liu, S. Sudek, S. Allen, C. Anderson, D. H. Sherman, and M. Haygood. 2004. BryA: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina*. *Chem. Biol.* **11**:1543–1552.
- Imsiecke, G., M. Custodio, R. Borojevic, R. Steffen, M. A. Moustafa, and W. E. Muller. 1995. Genome size and chromosomes in marine sponges [*Suberites domuncula*, *Geodia cydonium*]. *Cell Biol. Int.* **19**:995–1000.
- Jimeno, J., J. A. Lopez-Martin, A. Ruiz-Casado, M. A. Izquierdo, P. J. Scheuer, and K. Rinehart. 2004. Progress in the clinical development of new marine-derived anticancer compounds. *Anticancer Drugs* **15**:321–329.
- Kato, Y., N. Fusetani, S. Matsunaga, and K. Hashimoto. 1988. Calyculins, potent antitumor metabolites from the marine sponge *Discodermia calyx*: biological activities. *Drugs Exp. Clin. Res.* **14**:723–728.
- Kooistra, W. H., M. De Stefano, D. G. Mann, and L. K. Medlin. 2003. The phylogeny of the diatoms. *Prog. Mol. Subcell. Biol.* **33**:59–97.
- Lopez, J. V., P. J. McCarthy, K. E. Janda, R. Willoughby, and S. A. Pomponi. 1999. Molecular techniques reveal wide phylogenetic diversity of heterotrophic microbes associated with *Discodermia* spp. (Porifera: demospongia). *Mem. Queensland Museum* **44**:329–341.
- Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W.

- Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363–1371.
28. Manz, W., G. Arp, G. Schumann-Kindel, U. Szewzyk, and J. Reitner. 2000. Widefield deconvolution epifluorescence microscopy combined with fluorescence in situ hybridization reveals the spatial arrangement of bacteria in sponge tissue. *J. Microbiol. Methods* **40**:125–134.
29. Matsunaga, S., N. Fusetani, and S. Konosu. 1985. Bioactive marine metabolites. IV. Isolation and the amino acid composition of discodermin A, an antimicrobial peptide, from the marine sponge *Discodermia kiiensis*. *J. Nat. Prod.* **48**:236–241.
30. Minnikin, D. E., L. Kremer, L. G. Dover, and G. S. Besra. 2002. The methyl-branched fortifications of *Mycobacterium tuberculosis*. *Chem. Biol.* **9**:545–553.
31. Mirsky, A. E., and H. Ris. 1951. The desoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Physiol.* **34**:451–462.
32. Moffitt, M. C., and B. A. Neilan. 2003. Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *J. Mol. Evol.* **56**:446–457.
33. Newman, D. J., G. M. Cragg, and K. M. Snader. 2000. The influence of natural products upon drug discovery. *Nat. Prod. Rep.* **17**:215–234.
34. Piel, J. 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. USA* **99**:14002–14007.
35. Piel, J., D. Hui, N. Fusetani, and S. Matsunaga. 2004. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ. Microbiol.* **6**:921–927.
36. Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, and S. Matsunaga. 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci. USA* **101**:16222–16227.
37. Piel, J., G. Wen, M. Platzer, and D. Hui. 2004. Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *ChemBiochem* **5**:93–98.
38. Proksch, P., R. A. Edrada, and R. Ebel. 2002. Drugs from the seas: current status and microbiological implications. *Appl. Microbiol. Biotechnol.* **59**:125–134.
39. Rascher, A., Z. Hu, N. Viswanathan, A. Schirmer, R. Reid, W. C. Nierman, M. Lewis, and C. R. Hutchinson. 2003. Cloning and characterization of a gene cluster for geldanamycin production in *Streptomyces hygroscopicus* NRRL 3602. *FEMS Microbiol. Lett.* **218**:223–230.
40. Reid, R., M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson, and R. McDaniel. 2003. A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**:72–79.
41. Rezanka, T., and V. M. Dembitsky. 2002. Multibranched polyunsaturated and very-long-chain fatty acids of freshwater Israeli sponges. *J. Nat. Prod.* **65**:709–713.
42. Salomon, C. E., T. Deerinck, M. H. Ellisman, and D. J. Faulkner. 2001. The cellular localization of dercitamide in the Palauan sponge *Oceanapia sagittaria*. *Mar. Biol.* **139**:313–319.
43. Sambrook, J., and D. W. Russell. 2001. *Molecular Cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
44. Schleper, C., E. F. DeLong, C. M. Preston, R. A. Feldman, K. Y. Wu, and R. V. Swanson. 1998. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **180**:5003–5009.
45. Schloss, P. D., and J. Handelsman. 2003. Biotechnological prospects from metagenomics. *Curr. Opin. Biotechnol.* **14**:303–310.
46. Schmidt, E. W., A. Y. Obratsova, S. K. Davidson, D. J. Faulkner, and M. G. Haygood. 2000. Identification of the antifungal peptide-containing symbiont of the marine sponge *Theonella swinhoei* as a novel  $\delta$ -proteobacterium, “*Candidatus* Entotheonella palauensis”. *Mar. Biol.* **136**:969–977.
47. Seack, J., M. Kruse, I. M. Mueller, and W. E. G. Mueller. 1999. Promoter and exon-intron structure of protein kinase C gene from the marine sponge *Geodia cydonium*: evolutionary considerations and promoter activity. *Biochim. Biophys. Acta* **1444**:241–253.
48. Silakowski, B., H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blocker, G. Hoffe, S. Beyer, and R. Muller. 1999. New lessons for combinatorial biosynthesis from myxobacteria: the myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3–1. *J. Biol. Chem.* **274**:37391–37399.
49. Staunton, J., and K. J. Weissman. 2001. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**:380–416.
50. Trefzger, A., S. Pelzer, J. Schimana, S. Stockert, C. Bihlmaier, H. P. Fiedler, K. Welzel, A. Vente, and A. Bechthold. 2002. Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob. Agents Chemother.* **46**:1174–1182.
51. Turgay, K., and M. A. Marahiel. 1994. A general approach for identifying and cloning peptide synthetase genes. *Peptide Res.* **7**:238–241.
52. Unson, M. D., and D. J. Faulkner. 1993. Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia* **49**:349–353.
53. Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**:66–74.
54. Voget, S., C. Leggewie, A. Uesbeck, C. Raasch, K. E. Jaeger, and W. R. Streit. 2003. Prospecting for novel biocatalysts in a soil metagenome. *Appl. Environ. Microbiol.* **69**:6235–6242.
55. Webster, N. S., and R. T. Hill. 2001. The culturable community of the Great Barrier Reef sponge *Rhopaloeides odorabile* is dominated by an  $\alpha$ -proteobacterium. *Mar. Biol.* **138**:843–851.
56. Webster, N. S., A. P. Negri, M. M. Munro, and C. N. Battershill. 2004. Diverse microbial communities inhabit Antarctic sponges. *Environ. Microbiol.* **6**:288–300.
57. Webster, N. S., K. J. Wilson, L. L. Blackall, and R. T. Hill. 2001. Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile*. *Appl. Environ. Microbiol.* **67**:434–444.